AD_____

GRANT NUMBER DAMD17-94-J-4332

TITLE: Statistical Methods for Analyzing Time-Dependent Events in Breast Cancer Chemoprevention Studies

PRINCIPAL INVESTIGATOR: George Y.C. Wong, Ph.D.

CONTRACTING ORGANIZATION: Strang-Cornell Cancer Research Laboratory
New York, New York 10021

REPORT DATE: October 1996

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1996	Annual (30 Sep 95 - 29 Sep 96)

4. TITLE AND SUBTITLE
Statistical Methods for Analyzing Time-Dependent Events in Breast Cancer Chemoprevention Studies

5. FUNDING NUMBERS
DAMD17-94-J-4332

6. AUTHOR(S)
George Y.C. Wong, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Strang-Cornell Cancer Research Laboratory
New York, New York 10021

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

The overall aim of our research proposal is the statistical inference of nonparametric estimates, the redistribution-to-the-inside estimator (RTIE) and the generalized maximum likelihood estimator (GMLE), for the survival function, where the survival time is subject to interval censoring. The GMLE is the standard optimal procedure in survival analysis. However, a closed form expression for this estimator has not been derived, and the asymptotic distribution theory for it has been very little known (see Groeneboom and Wellner [1]). In our original proposal, we created the RTIE, which has a closed form expression, and which has been shown by us to be the GMLE under certain conditions. Working on the theory of the RTIE has provided us with important clues to the asymptotic theory concerning the GMLE. Our research efforts in the second year have focused on attacking the asymptotic distribution of the GMLE under the assumption that the censoring random vector is discrete. Under such an assumption, we have successfully established the asymptotic properties of the GMLE as well as those of the RTIE.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer, Interval Censorship, Asymptotic Normality and Efficiency	13
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

George Wang oct 28, 1996

PI - Signature Date

A. TABLE OF CONTENTS

B. INTRODUCTION

In clinical follow-up studies, subjects are monitored at regular time intervals for a physical condition. It is often the case that an event under observation can take place in between two successive visits, and it may not be possible for the subject to know the time to such an event exactly. For example, consider the situation in which a group of women at high risk for breast cancer is asked to take a chemopreventive substance for a fixed time period. At the end of the period, each participating woman is required to submit a blood or urine sample at regular intervals in order to monitor the level of a validated intermediate biomarker. Let X denote the time from cessation of use of the agent to the loss of its protective effect, quantified as a return to baseline value of the biomarker. If a woman submits a sample for assay on a daily basis, the value of X can be observed exactly, unless the protective effect is still present by the time the study is terminated so that X is right censored in the usual sense of survival analysis. In practice, however, the follow-up interval can be a week or longer; therefore the exact value of X is generally unknown but is known to lie between the time points L and R, where L is the number of days from cessation of agent intake to the last time the sample was assayed and the protective effect was still present, and R is the number of days from cessation of agent intake to the most recent time the sample was assayed. If the protective effect is still present, then R takes the value infinity. In any case, when the value of X is only known to lie between (L, R), we say that X is censored in the interval (L, R). Therefore the observed data consist of either censoring intervals (L, R) or exact observations $X = L = R$.

We consider nonparametric estimation of the distribution function $F(t)$ of a real-valued random variable X (or its survival function $S(t) = 1 - F(t)$, where $F(t) = \mathrm{P}\{X \leq t\}$), when the sample data are incomplete due to restricted observation brought about by interval censoring.

At present, there are only two estimation procedures of S for interval-censored data that are generalized maximum likelihood estimates (GMLE) in the sense of Kiefer and Wolfowitz [2]. The first one is due to Peto [3] and makes use of the Newton-Ralphon algorithm. The second is due to Turnbull [4] and makes use of a self-consistent algorithm. A solution to the latter algorithm is also called a self-consistent estimator (SCE) of S. In each case, there is no closed form expression for the estimator.

In the first year of our research, we have focused our attention on interval-censored data that satisfy a condition which we call DI condition: data $\{L_1, R_1\}, ..., \{L_n, R_n\}$ are said to satisfy DI condition if given any two censoring intervals, (L_i, R_i) and (L_j, R_j), either they are disjoint or one is a subset of the other. In a clinical study in which every subject has the same follow-up schedule, say at time point a_1, a_2, ..., a_k, then $\{L, R\} = \{0, a_1\}$, or $\{a_i, a_{i+1}\}$ or $\{a_i, \infty\}$, and hence such interval-censoring data will satisfy Condition DI.

Under the DI interval-censorship model, we have extended Efron's [5] redistribution -to-the-right idea for right-censored data and proposed a redistribution-to-the-inside (RTI) method to yield a nonparametric estimator of $S(t)$ which we call redistribution-to-the-inside estimator (RTIE). Such an estimate has a closed form expression and can be quickly calculated for interval-censored data of any size. The availability of an explicit expression for the RTIE has enabled us to show that it is the GMLE under the DI condition, and to establish asymptotic properties of the RTIE.

More often than not, interval-censored data do not satisfy the DI condition. In a clinical follow-up situation, for example, a patient may miss a particular appointment. Therefore, it is important to consider asymptotic inferences under a more general condition of interval censorship. Interval-censored data arise also quite naturally in medical follow-up studies or in industrial life-testing. A general interval censorship model can be described as follows: Suppose the survival time X has a distribution function F. What we really observe is an interval I, possibly a singleton set. If $I = [X, X]$, we have an exact observation; otherwise, we only know that X lies in the interval I, that is, the observation is interval censored. An observation is called *right censored* if I has a left endpoint ∞, *left censored* if I has a right endpoint 0, *exact* if I is a singleton set and *strictly interval censored* if the interval I is none of the above.

There are 4 typical situations in which interval-censored data can occur.

Case 2 interval-censored data (C2 data) consist of right-, left- and strictly interval-censored but not exact observations. Finkelstein and Wolfe [6] presented a set of case 2 interval-censored data in comparing two different treatments for breast cancer patients. The censoring intervals (in months) arose in the follow-up studies for patients treated with radiotherapy and chemotherapy. The failure time is the time until cosmetic deterioration, as determined by the appearance of breast retraction.

Partially interval-censored data (PIC data) consist of C2 data and exact observations. Yu, Li and Wong [7] presented a set of PIC data as follows.

Example 1. Three hundred and seventy-four women with stages I - III unilateral invasive breast cancer surgically treated on the Breast Service of Memorial Sloan-Kettering Cancer Center between 1985 and 1990 were followed for relapse. The median follow-up duration was 46 months. Relapse time was given by the time interval between surgery and the initial relapse. For a relapsed patient who was followed closely (for instance, during the initial follow-up period after surgery), an exact value for the relapse time could be meaningfully assessed. Otherwise, a relapse time between two successive follow-up visits would have to be regarded as interval censored. If a patient did not relapse towards the end of the study, then her relapse time was right censored. Of the 374 relapse times, 300 were right censored, 53 were interval censored, and 21 were observed exactly.

Doubly-censored data (DC data) consist of right-, left-censored and exact observations. Examples of DC data can be found in [8].

Case 1 interval-censored data (C1 data)) consist of right-censored and left-censored observations. Examples of C1 data can be found in [9] and [10].

Four different interval censorship models have been proposed corresponding to the four different types of data. They are the C2 model, the mixture interval censorship model (MIC model), the DC model and the C1 model. Only the C2 and the MIC models involve strictly interval-censored observations.

The GMLE for interval-censored data, is a distribution that maximizes the likelihood function (Kiefer & Wolfowitz [2]). The GMLE was derived via a numerical method by Peto [3] and Turnbull [4], and they conjectured that the GMLE has an asymptotic normal distribution. However, Groeneboom and Wellner [1] conjectured that it does not have the asymptotic normal distribution. So far, the asymptotic distribution of the GMLE of F has not been established for data involving strictly interval-censored observations (see, e.g.,

Groeneboom and Wellner [1]). Thus, in the research where interval-censored data occur, the current practice is to treat the strictly interval-censored data as right-censored data and to apply the Kaplan-Meier estimator. The asymptotic properties of the latter estimator have been well understood. However, this practice inevitably introduces biases in the statistical analysis.

To study the asymptotic properties of the GMLE, we make the following assumptions:

(AS1) The censoring distribution is discrete but the survival distribution is arbitrary.

(AS2) The censoring distribution has a support set of finitely many points, but the survival distribution is arbitrary.

In our second year, we have accomplished several important tasks for the GMLE under both DI and non-DI conditions:

1. Under the C1 model or the C2 model, we have proved the important result that the GMLE is strongly consistent under assumption (AS1)
2. Under the C1 model we have proved the important result that the GMLE is asymptotically normal and efficient under Assumption (AS1).
3. Under the C2 model we have proved the important result that the GMLE is asymptotically normal and efficient under assumption (AS2)
4. We proposed the MIC model for the PIC data.
5. Under the MIC model we have proved the important result that the SCE and the GMLE are strongly consistent under Assumption (AS1).
6. Under the MIC model we have proved the important result that the SCE and the GMLE are asymptotically normal and efficient under Assumption (AS1).

Four completed manuscripts ([7], [11], [12] and [13]), pertaining to results 1 thorough 6, have been submitted to peer-reviewed statistical journals. We are still preparing the fifth paper [14], pertaining to results 1 and 3. We presented some of our results at the Sydney International Statistical Congress, July 8-12, 1996, and at the Joint Statistical Meetings: Institute of Mathematical Statistics, American Statistical Association and International Biometric Society, August 4-8 Chicago.

C. BODY
Main Results

C.1. C2 Model.

By Assumption (AS1), there are only countably many (y_i, z_i)'s, we can assume that they are $\{a_1, a_2, ...\}$.

Theorem 1. *Under the C2 model and Assumption (AS1), the GMLE $\hat{F}(x)$ converges to $F(x)$ a.s. for all $x = a_i$, $i \geq 1$.*

Theorem 2. *Under the C2 model and Assumption (AS2), and suppose that there are altogether m points $a_1, ..., a_m$ and that $F(a_i) > F(a_{i-1})$ for $i = 2, ..., m$, we have, $\frac{\hat{F}(x)-F(x)}{\sigma} \xrightarrow{\mathcal{D}} N(0,1)$ as $n \to \infty$ for $x = a_i$, where σ^2 is given in Yu, Schick, Li and Wong [11].*

To see how close the approximation is to the theoretic results, we present numerical results in Table 1. The measure dF assigns the weight 0.2, 0.1, 0.25, 0.3 and 0.15 to the point 1, 3, 5, 7 and 9, respectively. The measure dG assigns the weight 0.4 and 0.6 to the point (2,6) and (4,8), respectively. In each simulation, the sample size of 800 was used. In the table, $\bar{\hat{F}}(x)$ stands for average of $\hat{F}$ with 1000 repetitions, $\mathrm{SD}(\hat{F}(x))$ for the sample standard deviation of $\hat{F}(x)$ and $\sigma(\hat{F}(x))$ for standard deviation of $\hat{F}(x)$ computed through formula given by Theorem 2.

Table 1. Standard Deviation of the GMLE

(x)	$F(x)$	$\bar{\hat{F}}(x)$	$SD(\hat{F}(x))$	$\sigma(\hat{F}(x))$
2	0.20	0.1996	0.0222	0.0224
4	0.30	0.3006	0.0207	0.0209
6	0.55	0.5512	0.0273	0.0278
8	0.85	0.8500	0.0165	0.0163

The sample SD's in the table match well with the values computed from the theoretic limits we have derived.

C.2. MIC Model.

The investigator proposed for PIC data the MIC model, which is a mixture of a C2 interval censorship model and a right censorship (RC) model (see Yu, Li and Wong [12]). The C2 model assumes that X is a non-negative random variable (failure time) with distribution function F and (Y, Z) is a non-negative random vector (censoring interval) with joint distribution function $G(u, v)$. It further assumes that $Y < Z$ with probability one (w.p.1), and that X and (Y, Z) are independent. The RC model assumes that there is a random censoring time T, with distribution function G_T, which is independent of X, and the information observed from the RC model is $(\min(X, T), I(X \leq T))$. We introduce a random variable, D, to distinguish failure times coming from the two models:

$$D = \begin{cases} 1 & \text{if the observation is from the RC model} \\ 0 & \text{if the observation is from the C2 model.} \end{cases}$$

Let $P\{D = 1\} = \pi$, where $0 < \pi \leq 1$. Formally, a PIC data point is regarded as an observation from the RC model w.p.π and from the C2 model w.p. $1 - \pi$.

To express observed PIC data as intervals, we introduce a notation $\lfloor L, R \rfloor$ defined as follows:

$$\lfloor L, R \rfloor = \begin{cases} [0, Y) & \text{if } D = 0 \text{ and } X < Y \\ [Y, Z) & \text{if } D = 0 \text{ and } Y \leq X < Z \\ [Z, \infty) & \text{if } D = 0 \text{ and } X \geq Z \\ (T, \infty) & \text{if } D = 1 \text{ and } X > T \\ [X, X] & \text{if } D = 1 \text{ and } X \leq T, \end{cases}$$

where $[X, X]$ is an exact observation. Let (L_i, R_i), $i = 1, 2, \ldots, n$ be a random sample from the random vector (L, R) with common joint distribution function $Q(l, r)$, and $\lfloor l, r \rfloor$ a realization of $\lfloor L, R \rfloor$. We say that the PIC data $\lfloor L, R \rfloor$ are from a *mixture* interval censorship model, called the MIC model.

Define $\tau = \sup\{t;\ P\{\min(X, T) \leq t\} < 1\}$ and $\tau_Z = \sup\{t;\ P\{Z \leq t\} < 1\}$. We assume that $\tau \geq \tau_Z$, which is imposed throughout the paper. This assumption is reasonable since under the RC model $[0, \tau]$ represents the whole time period of a follow-up study.

Define $\mathcal{O}_o = \{x;\ \mathrm{P}(X \text{ is not censored} | X = x) > 0\}$. Let $\mathcal{O}_c = \cap_{(l,r);\ r=\infty} \lfloor l, r \rfloor$, the intersection of all observed intervals with right endpoint infinity, and $\mathcal{O} = [0, \infty) \setminus \mathcal{O}_c$, where "\" is the set minus. For PIC data, it can be shown that $\mathcal{O} \subset [0, \tau]$. Whether $\mathcal{O} = [0, \tau]$ or not depends on F, G and G_T. To take the right endpoint τ into account, recall that under the RC model the strong consistency of the Kaplan-Meier estimator at τ requires either $F(\tau-) = 1$ or $P\{T = \tau\} > 0$ (cf. Yu and Li [15] p.416). Since the MIC model includes the RC model as a special case, a similar assumption is needed and is given as follows.

(AS3) Either $\mathrm{P}\{X \in \mathcal{O}\} = 1$ or $\mathrm{P}\{L = \tau\} > 0$.

Theorem 3. *Under (AS1) and (AS3), the SCE $\hat{F}(x)$ satisfies that*

$\lim_{n\to\infty} \sup_{x \in \mathcal{O}} |\hat{F}(x) - F(x)| = 0$ *a.s.*

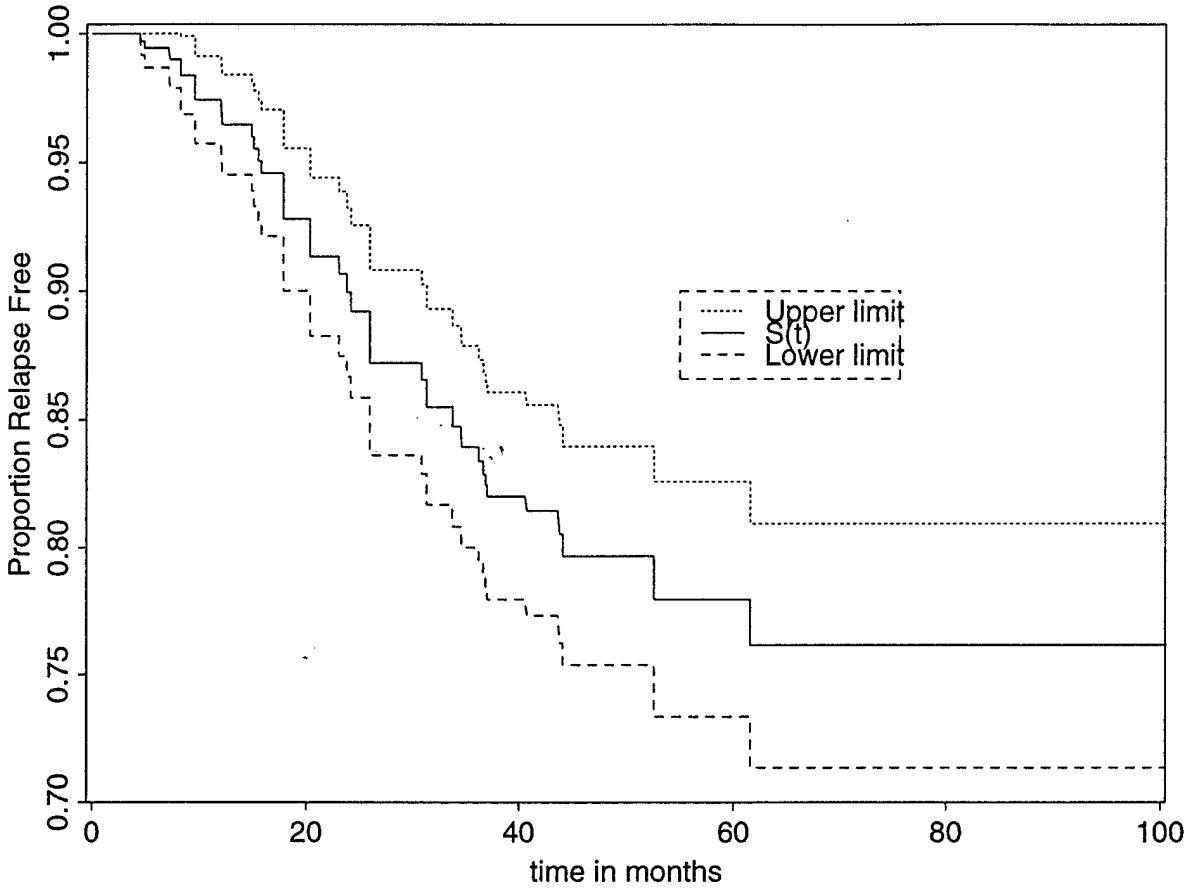
To establish asymptotic normality for the SCE, we need an additional assumption on the distribution function, namely,

(AS4) $\mathrm{P}\{X \in I_i \cap I_j\} > 0$ for any two realizations, I_i and I_j, of $\lfloor L, R \rfloor$, provided $I_i \cap I_j \neq \emptyset$.

Theorem 4. *Under Assumptions (AS1), (AS3) and (AS4), the SCE $\hat{F}(x)$ satisfies that for $x \in \mathcal{O}$, $\frac{\hat{F}(x) - F(x)}{\hat{\sigma}} \xrightarrow{\mathcal{D}} N(0, 1)$ as $n \to \infty$, where the notations are the same as in Theorem 2.*

We apply Theorem 4 to the breast cancer data in Example 1 to obtain the SCE and its asymptotic variance for the survival function $S(t)$, which represents the proportion of women who were relapse free at time t. Figure 1 gives the survival plot together with the 95% asymptotic confidence bands.

Fig.1. Self-Consistent Estimate for Breast Cancer Data



C.3. C1 Model.

By Assumption (AS1), there are only countably many (y_i, z_i)'s, WLOG, we can assume that they are $\{a_1, a_2, ...\}$.

Theorem 5. *Under the C1 model and Assumption (AS1), the GMLE $\hat{F}(x)$ converges to $F(x)$ a.s. for all $x = a_i$, $i \geq 1$.*

Theorem 6. *Under the C1 model and Assumption (AS1), and suppose that $F(z) > F(x) > F(y)$ for $z, x, y \in \{a_i\}_{i \geq 1}$ and $z < x < y$; and there is no other $a_i \in (z, y)$ other than x. Then we have, $\frac{\hat{F}(x) - F(x)}{F(x)[1-F(x)]g(x)} \xrightarrow{\mathcal{D}} N(0,1)$ as $n \to \infty$ for $x = a_i$, where $g(x) = P\{X = x\}$.*

D. CONCLUSIONS

As we point out in INTRODUCTION, interval-censored data are commonly encounted in cancer follow-up studies and there has been a lack of asymptotic estimation procedures for the survival function. In our second year of research, we have derived the asymptotic distribution for the GMLE under Assumption (AS1) or (AS2). In the BODY section, we have used our asymptotic results for the MIC model to produce the survival curve and its 95% confidence band plots for overall relapse free survival for interval-censored data from 374 women with stages I, II and III breast cancer after treatment by surgery.

Our immediate research goals for the third year are to extend the results established here to the case that the distribution functions are more general than those in assumptions (AS1) and (AS2). Specifically, we will extend the method to obtain the asymptotic distribution of the GMLE under (AS1) or (AS2) to the general case in which the distribution functions are arbitrary. We expect these extensions to be statistically fairly challenging.

E. REFERENCES

[1] Groeneboom, P. and Wellner, J. A. (1992). Information bounds and nonparametric maximum likelihood estimation. *Birkhäuser Verlag, Basel.*

[2] Kiefer, J and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* 27, 887-906.

[3] Peto, R. (1973). Experimental survival curves for interval-censored data. *Appl. Statist.* 22, 86-91.

[4] Turnbull, B. W. (1976). The empirical distribution function with arbitrary grouped, censored and truncated data. *J. Roy. Statist. Soc. Ser.* B, 38, 290-295.

[5] Efron, B (1967). The two sample problem with censored data. *Fifth Berkeley Symposium on Mathematical Statistics.* University of California Press, 831-853.

[6] Finkelstein, D.M. and Wolfe, R.A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics*, 41, 933-945.

[7] Yu, Q., Li, L. and Wong, G. Y. C. (1996a). On consistency of the self-consistent estimator of survival functions with interval censored data. Submitted to *Statistica Sinica.*

[8] Leiderman, P.H., Babu, D., Kagia, J., Kraemer, H.C. and Leiderman, G.F. (1973). African infant precocity and some social influences during the first year. *Nature*, 242, 247-249.

[9] Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T. and Silverman, E. (1955). An empirical distribution function for sampling incomplete information. *Ann. Math. Statist.* 26, 641-647.

[10] Keiding. N. (1991) Age-specific incidence and prevalence: A statistical perspective (with discussion) *JRSS*, A. 154, 371-412.

[11] Yu, Q., Schick, A., Li, L. and Wong, G. (1996a). Estimation of a survival function with case 1 interval-censored data. Submitted to *Biometrika.*

[12] Yu, Q., Li, L. and Wong, G. Y. C. (1996b). A model for partially-interval-censored data. Submitted to *Ann. Statist.*

[13] Yu, Q., Li, L. and Wong, G. Y. C. (1996b). Variance of the self-consistent estimator of survival functions with interval censored data. Submitted to *Sankhya.*

[14] Yu, Q., Schick, A., Li, L. and Wong, G. (1996c). Asymptotic properties of the GMLE of a survival function with case 2 interval-censored data. To be submitted to *Statist. & Prob. Let.*).

[15] Yu, Q. and Li, L. (1994). On the strong uniform consistency of the product limit estimator. *Sankhyā.* A. Vol. 56.

F. APPENDICES

1. A list of paper submitted:

[1] Yu, Q. and Wong, G. Y. C. (1994). Estimation of a survival function with interval-censored data under the DI model. (Submitted to *Ann. Inst. Statist. Math.*).

[2] Yu, Q. and Wong, G. Y. C. (1994). Strong consistency of the generalized mle of a survival function under the DI model. (Under preparation).

[3] Yu, Q., Li, L. and Wong, G. Y. C. (1996a). On consistency of the self-consistent estimator of survival functions with interval censored data. Submitted to *Statistica Sinica.*

[4] Yu, Q., Schick, A., Li, L. and Wong, G. (1996a). Estimation of a survival function with case 1 interval-censored data. Submitted to *Biometrika.*

[5] Yu, Q., Li, L. and Wong, G. Y. C. (1996b). A model for partially-interval-censored data. Submitted to *Ann. Statistics.*

[6] Yu, Q., Li, L. and Wong, G. Y. C. (1996b). Variance of the self-consistent estimator of survival functions with interval censored data. Submitted to *Sankhya.*

[7] Yu, Q., Schick, A., Li, L. and Wong, G. (1996c). Asymptotic properties of the GMLE of a survival function with case 2 interval-censored data. To be submitted to *Statist. & Prob. Let.*).

[9] Yu, Q. and Li, L. (1994). On the strong uniform consistency of the product limit estimator. *Sankhyā.* A. Vol. 56.

2. A list of conferences attended:

Two meetings.

(1) *Institute of Mathematical Statistics Meeting no. 246 8-12 July 1996 Sydney, N.S.W.*

TITLE: Variance of the MLE of a Survival Function with Interval Censored Data

Qiqing /Yu , Linxiong /Li and George Y. C. /Wong

SUNY at Binghamton, University of New Orleans and Strang Cancer Preventive Institute

ABSTRACT: Interval-censored data consist of n pairs of observations (l_i, r_i), $i = 1, ..., n$, where $l_i \leq r_i$. We either observe the exact survival time X if $l_i = r_i$ or only know $X \in (l_i, r_i)$ otherwise. We established the asymptotic normality of the nonparametric MLE of a survival function $S(t)$ $(= P(X > t)$ with such interval-censored data and present an estimate of the asymptotic variance of the MLE. We show that the convergence rate in distribution is in $\sqrt{n}$. Simulation study also supports our result. An application to the cancer research is presented.

[Corresponding author: Qiqing Yu, Math department, SUNY at Binghamton, NY 13902, qyu@math.binghamton.edu]

Paper presented in person, contributed paper.

(2) *Institute of Mathematical Statistics Meeting no. 247 4-8 August 1996 Chicago, Illinois*

TITLE: ESTIMATION OF A SURVIVAL FUNCTION WITH CASE 1 INTERVAL -CENSORED DATA

Qiqing /Yu , Anton /Schick, Linxiong /Li and George Y. C. /Wong
SUNY at Binghamton, University of New Orleans and Strang Cancer Preventive Institute

ABSTRACT: Case 1 interval censored data consists of either right-censored data or left-censored data but not exact observations. Let $F(x)$ and $G(y)$ be the distribution functions of the survival time X and censoring time Y, respectively. Groeneboom and Wellner (1992, p.100) (G&W) establish the consistency and asymptotic distribution of the MLE of F under the assumption that $F'(x)$ and $G'(y)$ are both positive and continuous. Under the assumption that X *is arbitrary, but* Y *takes on finitely many values*, we establish the consistency, asymptotic normality and efficiency of the MLE of $F(y)$ at the observations, y, of Y, and present a consistent estimate of the asymptotic variance of the MLE. The convergence rate in distribution is $n^{1/3}$ under G&W's assumption, but it is $\sqrt{n}$ under our assumption. Simulation results indicates that the sample variance is very close to the theoretical value of the asymptotic variance given in our paper, even for a sample size of 100.
[Corresponding author: Qiqing Yu, Math department, SUNY at Binghamton, NY 13902, qyu@math.binghamton.edu]

Paper presented in person, contributed paper.